

# DÉVELOPPEMENT D'UN SYSTÈME DE CLASSIFICATION AUTOMATIQUE DE L'ACTIVITÉ PHYSIQUE

Mémoire présenté comme exigence partielle  
de la maîtrise ès sciences (technologie de l'information)

Par Tariq Abdessettar

Février 2018



<http://r-libre.teluq.ca/1397>



Cette licence [Creative Commons](https://creativecommons.org/licenses/by-nc-nd/4.0/) signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

# Table des matières

Remerciement.....	1
Résumé.....	2
Abstract.....	3
Chapitre 1 : Introduction.....	4
1.1. Objectifs et hypothèses de recherche.....	5
1.2. Plan du mémoire.....	6
Chapitre 2 : Revue de la littérature.....	7
2.1. Les capteurs utilisés dans les téléphones intelligents.....	7
A. Accéléromètre.....	8
B. Gyroscope.....	9
C. Le capteur de gravité.....	10
D. Magnétomètre.....	11
2.2. Principe de la classification de formes.....	12
A. Prétraitement.....	13
B. Extraction et sélection des caractéristiques.....	14
C. Classification.....	14
D. Évaluation et validation d'un classificateur.....	15
2.3. Techniques de classification.....	18
La classification supervisée.....	18
La classification non supervisée.....	20

2. 4. Classification de l'activité physique.....	21
Chapitre 3 : Système de classification de l'activité physique.....	27
3.1. Méthodologie.....	27
A. Conception de la base de données.....	28
B. Développement d'interfaces.....	30
C. Collecte de données .....	31
D. Système de classification.....	34
3.5. Résultats expérimentaux : Tests et validation.....	36
CONCLUSION.....	42
RÉFÉRENCES BIBLIOGRAPHIQUES.....	44

# Remerciement

La réalisation de ce mémoire a été possible grâce à plusieurs personnes. Je voudrais les remercier et témoigner toute ma reconnaissance.

Je voudrais tout d'abord adresser toute ma gratitude à la directrice de ce mémoire, Mme Neila Mezghani, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion.

Je remercie M. Youssef Ouakrim, assistant de recherche au laboratoire de recherche en imagerie et orthopédie LIO pour son aide et ses conseils ainsi que tout le personnel du Laboratoire de recherche en imagerie et orthopédie LIO.

Je remercie spécialement M. Saad Abdessettar pour sa patience, ses conseils et son soutien durant toute ma formation.

Je remercie également M. Karim Ait-Seddik pour sa collaboration.

Je voudrais exprimer ma reconnaissance envers les amis et collègues qui m'ont apporté leur appui moral et intellectuel tout au long de ma démarche. Un grand merci à tout le personnel des deux laboratoires de recherche LIO-CRCHUM et LICEF et au corps professoral de la TÉLUQ.

Finalement, je tiens à remercier mes parents et mes frères pour leurs encouragements.

Merci à tous.

# Résumé

La santé mobile (m-santé) couvre «...les pratiques médicales et de santé publique reposant sur des dispositifs mobiles tels que les téléphones portables, les systèmes de surveillance des patients, les assistants numériques personnels et autres appareils sans fil ». Elle englobe également les applications concernant le mode de vie et le bien-être qui peuvent se connecter à des dispositifs médicaux ou capteurs ainsi que les systèmes de conseil personnalisés, les informations de santé et rappels de prise de médicament envoyés par SMS et la télémédecine pratiquée par communication sans fil.

La santé mobile est un nouveau domaine d'activité dont le développement rapide peut contribuer à faire évoluer les soins de santé et à accroître leur qualité et leur efficacité. Cette évolution est rendue possible grâce aux nouveaux appareils mobiles intelligents, tels que, les téléphones et les tablettes qui contiennent des capteurs (en anglais *sensors*) et qui permettent de collecter plusieurs types de données par exemple le mouvement (l'accélération, le capteur de gravité et le gyroscope), les capteurs de position (magnétomètre, proximité et GPS (Global Positioning System)) et les capteurs qui permettent d'enregistrer divers signaux physiologiques (signal cardiaque, signal respiratoire, la tension artérielle et la température corporelle).

Ce projet s'insère dans cette veine de recherche. Il a pour objectif de développer un système de classification automatique de l'activité physique de l'utilisateur d'un terminal mobile en utilisant les données captées par le terminal. La majorité des études se sont basées sur l'accéléromètre afin de se servir des valeurs des trois composantes d'accélérations et/ou de l'accélération linéaire résultante des taux de classification qui varient entre 64,5% et 99,1% (Robertas Damasevicius et al [25]) selon le type de classificateur et le nombre d'activités physiques à classifier.

La classification de l'activité physique est une problématique de recherche importante en santé mobile. En effet, la classification de l'activité cardiaque et respiratoire dépend directement de l'activité physique d'où l'importance de reconnaître l'activité physique pour pouvoir détecter les anomalies cardiaques et respiratoires.

# Abstract

Mobile health (m-health) covers "medical practices and public health-based mobile devices such as mobile phones, patient monitoring systems, PDAs and other wireless devices." It also includes applications for lifestyle and well-being that can connect to medical devices or sensors and custom consulting systems, health information and taking medication reminders sent by SMS and télémédecine practiced by wireless communication.

Mobile health is a new area of activity whose rapid development can contribute to changing health care and increase quality and efficiency. This development was made possible thanks to new smart mobile devices, such as, phones and tablets that contain sensors and that collect several types of data, such as the movement (accelerometer, gravity and gyroscope), the position sensors (magnetometer, proximity and GPS (Global Positioning System)) and sensors that can record various physiological signals (heart, blood pressure, ...).

This project will insert in this research stream. It aims to develop an automatic classification system of physical activity of the user of a mobile terminal using the data captured by the terminal. The majority of the studies were based on the accelerometer to use the values of the three components of accelerations and / or the linear acceleration resulting from the classification rates which vary between 64.5% and 99.1% (Robertas Damasevicius et al [25]) according to the type of classifier and the number of physical activities to classify.

The classification of physical activity is an important research topic in mobile health. In fact the classification of cardiac activity and respiration depends directly on the physical activity and the importance of recognizing physical activity to detect cardiac and respiratory abnormalities.

# Chapitre 1

## Introduction

Les développements récents des technologies mobiles (par exemple, téléphones portables, systèmes de surveillance des patients, assistants numériques personnels) ont eu un impact extrêmement remarquable sur le comportement humain. En effet, ces appareils mobiles contiennent des capteurs qui peuvent jouer un rôle important pour améliorer la qualité de vie de leurs usagers, que ce soit dans le domaine de la santé, dans le domaine des finances ou dans la gestion de la vie quotidienne.

Dans le domaine de la santé, qui nous intéresse ici, l'utilisation des technologies mobiles est appelée « Santé mobile » (en anglais *mobile health* ou *m-health*). Ces technologies émergent tous les jours; ils offrent une multitude de fonctionnalités comme des plans d'entraînement physique, de l'information sur les bonnes habitudes et sur la nutrition ou encore des journaux de bord pour la gestion des maladies chroniques. Ces fonctionnalités utilisent souvent comme intrant des données provenant des capteurs de terminaux mobiles. Parmi ces capteurs, on trouve les capteurs de mouvement (l'accélération, le capteur de gravité et le gyroscope), les capteurs de position (magnétomètre, proximité et GPS (Global Positioning System)) et les capteurs qui permettent d'enregistrer divers signaux physiologiques (signal cardiaque, signal respiratoire, la tension artérielle et la température corporelle). Ainsi les technologies de santé mobile permettent de recueillir, à l'aide de capteurs et d'applications mobiles, un volume considérable de données médicales, physiologiques, sur le mode de vie, l'activité quotidienne et l'environnement. Ces données pourraient servir de base à des pratiques médicales et des activités de recherche reposant sur les faits, tout en permettant aux patients d'accéder à leurs informations de santé partout et à tout moment.



D'ailleurs la santé mobile devient un moyen complémentaire de dispenser des soins. Une récente étude de l'OMS [51] montre que, « dans les pays à revenu élevé, la santé mobile répond à l'impératif de réduire les dépenses de santé tandis que, dans les pays en développement, elle se justifie surtout par la nécessité d'accéder aux soins primaires. L'étude a également révélé que l'un des plus récents facteurs de progrès sanitaire est constitué par les systèmes qui permettent de dispenser des soins personnalisés, à l'aide de dispositifs portables, portatifs ou implantables, et confèrent aux patients un rôle plus actif (appelés systèmes de santé individuels). ».

L'objectif général de ce projet de recherche consiste à développer un système de classification automatique des activités physiques de l'utilisateur d'un appareil mobile. Ce système se basera sur les données enregistrées à partir des capteurs du terminal mobile pour identifier les classes d'activités.

La classification de l'activité physique occupe de plus en plus une position importante dans notre vie quotidienne. À titre d'exemple, un programme national de nutrition santé a été mis en place en France pour décrire le type de l'alimentation en fonction selon le type de l'activité physique pratiqué régulièrement. D'où le besoin de classer en continu cette activité. Un autre exemple, porte sur la télé réadaptation qui est un nouveau type de prestation de services à distance dont il est possible d'espérer qu'il pallie à une pénurie de services de réadaptation, et ce, à un coût avantageux pour l'ensemble de la société. Pour réaliser la télé réadaptation, il sera nécessaire de développer des systèmes de classification automatique de l'activité physique.

## **1.1. Objectifs et hypothèses de recherche**

L'objectif principal de ce mémoire est de développer un système de classification automatique des activités physiques. Pour y parvenir, nous avons défini les objectifs spécifiques suivants :

- 1- Développement d'une application mobile pour collecter les données des quatre capteurs (accéléromètre, gyroscope, gravité et magnétomètre).

2- Développement d'une application web pour l'affichage et le prétraitement des données enregistrées.

3- Conception du système de classification des données prétraitées afin d'identifier les différentes classes d'activités physiques.

Notre hypothèse de recherche est donc qu'il est possible de classifier les activités physiques en utilisant les données provenant d'un terminal mobile.

## **1.2. Plan du mémoire**

Ce mémoire sera organisé comme suit :

Après ce premier chapitre d'introduction, nous ferons dans le chapitre 2 une revue de la littérature qui couvrira les trois aspects suivant (1) les capteurs utilisés dans les téléphones intelligents (2) les techniques de classification de formes en général et (3) plus spécifiquement pour l'activité physique.

Dans le chapitre 3, nous décrivons le système de classification de l'activité physique que nous avons développée. Nous présenterons les différents modules à savoir, l'interface de collecte de données, l'extraction des caractéristiques et la classification. Nous détaillerons les différentes expériences qui ont abouti à la validation du système de classification.

Finalement, dans le chapitre 4, nous résumons les contributions apportées dans ce mémoire ainsi que les résultats obtenus et nous concluons avec les perspectives reliées aux travaux futurs.

# Chapitre 2

## Revue de la littérature

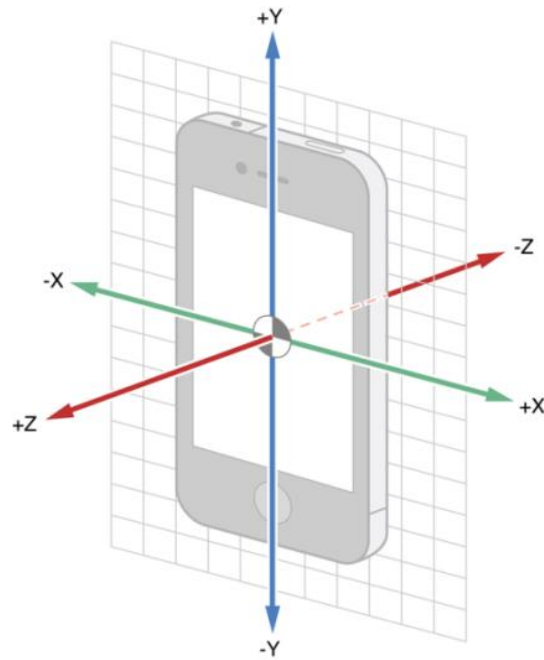
L'objectif de ce projet de recherche, rappelons-le, est le développement d'un système de classification automatique des activités physiques de l'utilisateur d'un appareil mobile. Ce système utilisera comme intrant les données enregistrées à partir des capteurs de l'appareil mobile. Nous proposons dans ce chapitre une revue de la littérature qui couvrira les trois aspects suivants. Le premier portera sur les capteurs utilisés dans les téléphones intelligents (Section 2.1). Le deuxième constituera un survol du principe général de classification de forme et des méthodes de classification de formes utilisées dans la littérature (Section 2.2). Le troisième aspect traitera, plus spécifiquement, les méthodes de classification de formes utilisées dans le cadre de la classification de l'activité physique (Section 2.3)

### **2.1. Les capteurs utilisés dans les téléphones intelligents**

La reconnaissance du mouvement humain à base de capteurs installés sur des terminaux mobiles a reçu beaucoup d'attention de la part des chercheurs ces dernières années. De ce fait, plusieurs capteurs ont été adaptés pour les mouvements à grande échelle pour améliorer la qualité de vie des personnes. Ces capteurs présentent deux avantages majeurs : (1) la portabilité par rapport aux méthodes basées sur la vision par ordinateur et de télésurveillance et (2) la reconnaissance des mouvements du corps entier et le contrôle de mouvement ce qui permet de les appliquer dans divers domaines, tels que la réadaptation médicale et des jeux interactifs.

La plupart des appareils mobiles, téléphones intelligents et tablettes sont munis de capteurs qui permettent de convertir des valeurs analogiques en des valeurs numériques afin de pouvoir lire les données collectées par un appareil mobile. Il existe plusieurs types de capteurs. Cependant, dans cette section nous nous sommes limités à ceux qui sont fréquemment intégrés dans les appareils mobiles à savoir, l'accéléromètre, le gyroscope, le capteur de gravité et le magnétomètre.

Les trois premiers capteurs permettent de détecter le mouvement et la rotation tandis que le magnétomètre joue le rôle d'une boussole grâce au champ magnétique. Les données enregistrées par ces capteurs sont tridimensionnelles, c'est-à-dire selon les trois axes X, Y et Z telles que illustrées dans la Figure 1.



**Figure 1 : Référentiels utilisés dans les terminaux mobiles**

### **A. Accéléromètre**

L'accéléromètre permet de mesurer l'accélération linéaire selon trois axes orthogonaux (X, Y et Z) ce qui permet de déduire les changements de vitesse et les changements de

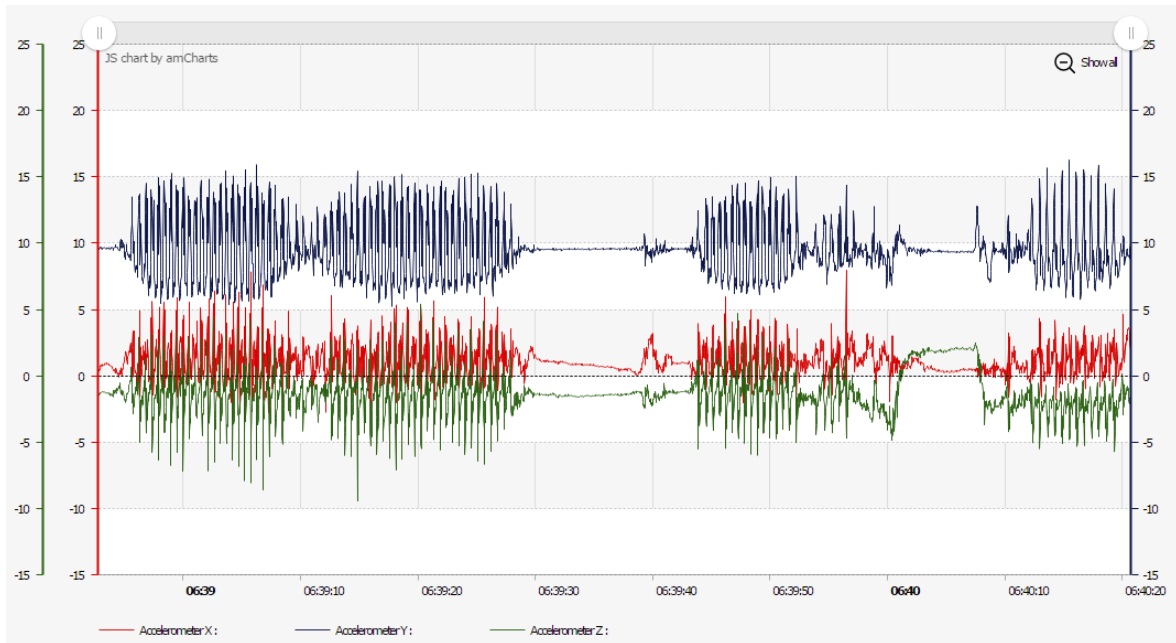
position. L'accéléromètre permet aussi de déterminer l'orientation de l'écran d'un téléphone intelligent (paysage et portrait) grâce à la force de gravité générée par la terre.

L'accélération linéaire est calculée en utilisant la formule suivante :

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad (1.1)$$

Où A est la valeur de l'accélération linéaire; elle est exprimée en  $m/s^2$ .  $A_x$ ,  $A_y$  et  $A_z$  désignent les valeurs des composantes de l'accélération selon les trois axes X, Y et Z.

La figure 2 montre des valeurs enregistrées par un accéléromètre selon les trois axes (composante X, composante Y, composante Z).



**Figure 2 : Signaux d'accélération collectés à partir de terminaux mobiles  
(En rouge  $A_x$  en bleu  $A_y$  et en vert  $A_z$  ).**

## **B. Gyroscope**

Le gyroscope permet de mesurer la position angulaire et donc la vitesse de rotation selon les trois axes (X, Y et Z). Dans le cas des appareils mobiles, le gyroscope est calibré pour

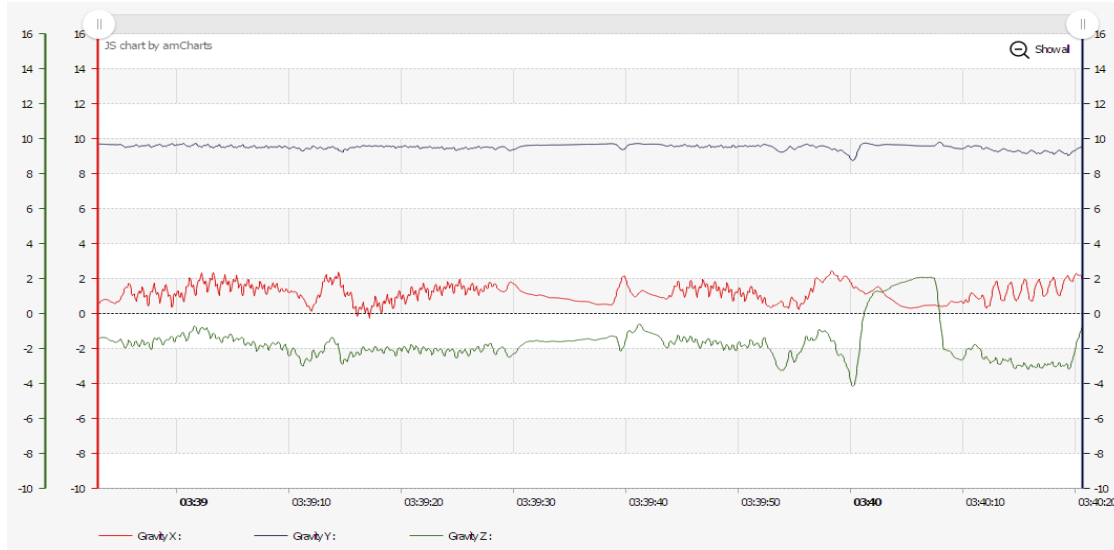
donner une lecture de zéro lorsque le dispositif est maintenu sur une surface horizontale. Tout changement dans l'orientation du dispositif est mesuré par le capteur gyroscopique. Un exemple d'utilisation du gyroscope est, lorsqu'on a une image visualisée dans un plan horizontal et que l'on change le plan vers celui correspondant à la verticale on remarque que l'image s'oriente et s'affiche dans le plan vertical. La figure 3 illustre un exemple de données enregistrées par le gyroscope.



**Figure 3 : Données collectées à partir d'un gyroscope**  
(En rouge  $A_x$  en bleu  $A_y$  et en vert  $A_z$  ).

### **C. Le capteur de gravité**

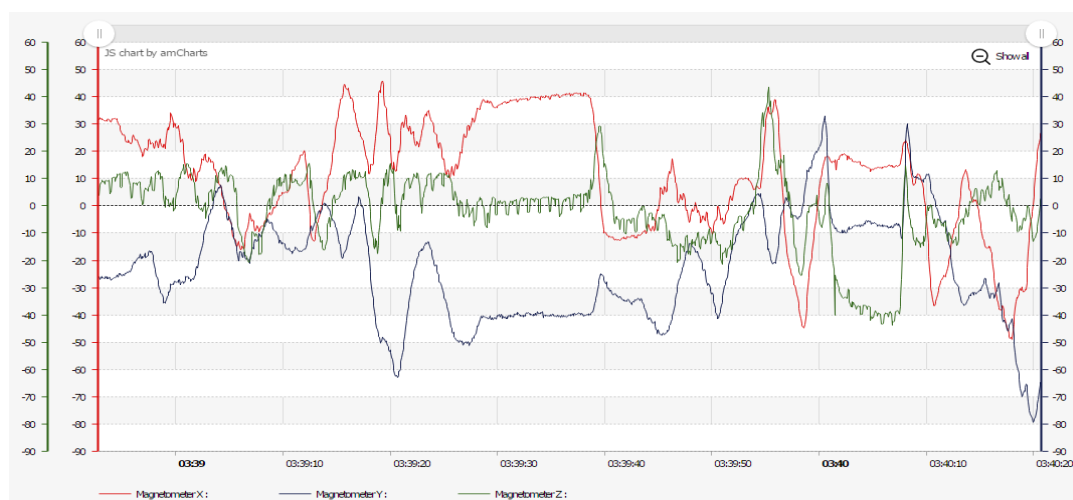
Le capteur de la gravité fonctionne selon le même principe que celui de l'accéléromètre sauf qu'il permet de mesurer, non pas l'accélération linéaire, mais la gravité par rapport à la terre. La figure 4 montre un exemple des valeurs enregistrées par un capteur de gravité.



**Figure 4 : Données collectées par le capteur de gravité.**  
(En rouge  $A_x$  en bleu  $A_y$  et en vert  $A_z$  ).

#### **D. Magnétomètre**

Le magnétomètre est un capteur qui permet de mesurer le champ magnétique par rapport à la terre, et ce, dans les trois axes (X, Y et Z). Il produit une tension proportionnelle à la force et la polarité du champ magnétique. La figure 5 montre un exemple des valeurs enregistrées par un magnétomètre.



**Figure 5 : Données collectées par le magnétomètre**  
(En rouge  $A_x$  en bleu  $A_y$  et en vert  $A_z$  ).

## **2.2. Principe de la classification de formes**

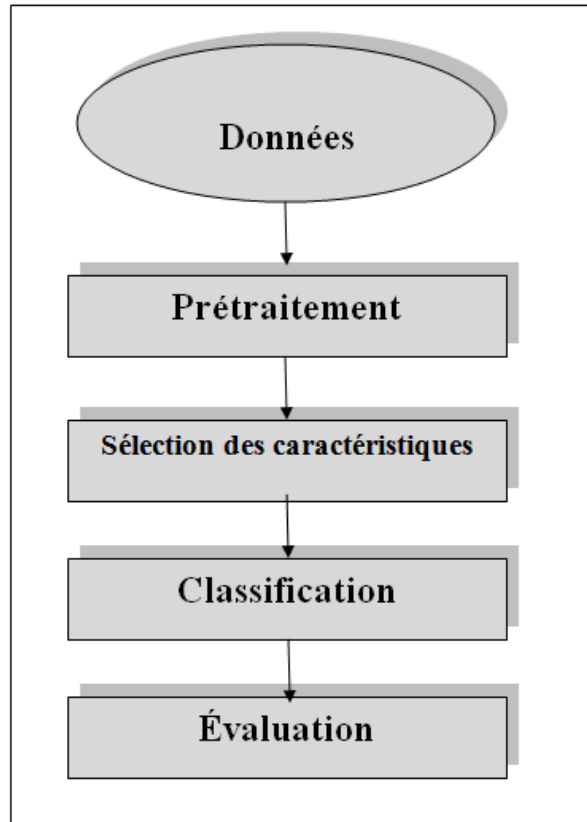
De nos jours, le volume des données collectées ne cesse d'augmenter exponentiellement et le critère du temps devient important. De ce fait, les techniques de classification automatique de données sont devenues très importantes pour faciliter l'exploitation de ces données. Plusieurs recherches et études ont été réalisées dans le domaine de classification de forme pour permettre de mieux comprendre et d'exploiter les données surtout dans le cas où ces données sont de volume important et que les critères de sélection sont moins visibles.

L'intégration de ces techniques dans la reconnaissance des activités physiques est devenue essentielle pour améliorer la qualité de vie humaine, et ce, en utilisant des technologies mobiles telles que les téléphones intelligents et les textiles intelligents qui contiennent des capteurs permettant de définir le type d'activité effectuée par une personne.

Le processus de la classification de formes commence par une première étape de prétraitement pour éliminer les différents bruits des données enregistrées. La deuxième étape consiste à définir les différentes caractéristiques à utiliser lors de la classification. Parmi ces caractéristiques, on trouve la moyenne, l'écart-type, la médiane (Yi He et Ye Li [50]; Oscar D et al [48]). L'étape suivante consiste à développer le classificateur pour reconnaître les différentes classes et la dernière étape permet d'évaluer et valider le système de classification.

Généralement, le diagramme en bloc d'un système de classification de formes suit la figure suivante qui illustre les étapes du processus de classification.





**Figure 6 : Processus de classification**

### **A. Prétraitement**

Le processus de classification est précédé par une phase de prétraitement qui consiste à réduire les bruits causés par les conditions d'acquisition des données et d'éliminer les distorsions et les données inutiles [8]. Le prétraitement consiste généralement à faire un filtrage pour éliminer le bruit et à normaliser les données afin d'uniformiser la dimension ou l'échelle.

Plusieurs techniques de filtrage existent dans la littérature parmi lesquels on retrouve, le filtre de *Kalman*. Le filtre de *Kalman* est un filtre à réponse impulsionnelle infinie. Ce filtre permet d'estimer l'état d'un processus tout en minimisant la moyenne de l'erreur quadratique. Le filtre de *Kalman* est très puissant parce qu'il prend en considération les estimations des états passés, actuels et futurs. Ce filtre est un outil important dans l'étude des systèmes dynamiques parce qu'il offre un cadre fondamental à l'estimation d'états.

Cependant, il présente certaines limitations. En effet, il présente des instabilités dues à des écarts de modèle et à des problèmes de calcul en précision finie. De plus, est confronté à des problèmes de divergence. Ces problèmes peuvent être limités en utilisant un filtrage avec une équation qui propage la racine carrée de la matrice de covariance.

## **B. Extraction et sélection des caractéristiques**

Le but de cette étape est d'identifier les caractéristiques qui permettent de discriminer les différentes classes. Dans le cadre de l'analyse de l'activité physique, Oscar et al [48] ont recensé deux types de caractéristiques : les caractéristiques basées sur la fréquence (caractéristiques fréquentielles) et les caractéristiques basées sur le temps (caractéristiques temporelles) telles que la moyenne, la médiane, l'écart-type, l'interquartile [50].

D'autres types de caractéristiques qui permettent une représentation globale peuvent être utilisés. En occurrence, l'analyse en composantes principales (ACP) est une méthode d'analyse de données très connue en statistique et dans les sciences expérimentales. Elle consiste à rechercher les directions de l'espace qui représentent le mieux les corrélations dans un ensemble de données. Ceci a pour objectifs de réduire la dimension des caractéristiques et de les visualiser afin d'analyser les corrélations entre ces données.

## **C. Classification**

La classification de formes est l'« Opération qui consiste à grouper formellement des éléments de configuration, des incidents, des problèmes ou des changements, en fonction de leur type » (Office québécois de la langue française, 2005). En d'autres termes, la classification de formes consiste à affecter un objet à une classe correspondante en se basant sur leurs caractéristiques.

Les méthodes de classification de formes peuvent être réparties en deux types. La classification supervisée et la classification non supervisée. Ces techniques seront détaillées à la section 2.3 de ce chapitre.

## **D. Évaluation et validation d'un classificateur**

L'évaluation d'un système de classification se base généralement sur un partage de la base de données en trois domaines aussi appelés sous-bases :

- Une **sous-base d'apprentissage** pour la conception du système de classification ;
- Une **sous-base de test** pour sélectionner l'architecture du système de classification ;
- Une **sous-base de validation** utilisée pour valider la pertinence de l'architecture choisie.

La distinction entre les deux premières bases est nécessaire pour ne pas biaiser les tests de classification en testant sur des échantillons ayant servi pour la conception. La 3e sous-base permet d'obtenir une validation indépendante des critères de sélection. Dans certains cas, la taille de la base de données ne permet pas d'avoir un partage en trois sous-bases ; dans ce cas on utilise uniquement un partage en deux : une sous-base d'apprentissage et une sous-base de tests.

Il existe différents types d'algorithmes pour réaliser le partage des données en données d'apprentissage et données de test. Parmi ces méthodes on trouve la validation par K-fold.

- **Validation par K-fold**

Le principe de cette méthode de validation consiste à diviser l'échantillon original en K échantillons de même taille. Puis on prend un échantillon pour procéder au test. On répète le processus jusqu'on atteint les K échantillons. En d'autres termes, on divise l'échantillon original en K échantillons, puis on sélectionne un des K échantillons comme ensemble de tests et les (K-1) autres échantillons constitueront l'ensemble d'apprentissages pour la conception du système de classification. À la fin, on prend la moyenne des résultats de test pour avoir un seul résultat.

- **Critères de validation**

L'évaluation d'un système de classification se base généralement sur la mesure du taux de classification. Pour avoir une meilleure interprétation des résultats, nous utilisons la matrice de confusion. Il s'agit d'une matrice carrée de taille  $[W \times W]$  où  $W$  représente le nombre de classes. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe prédite, alors que chaque ligne représente le nombre d'occurrences d'une classe réelle (ou de référence). La matrice de confusion offre la possibilité d'analyse du taux de bonne classification (ou d'erreur) par classe.

Les valeurs de la diagonale de la matrice représentent le nombre de formes correctement classifiées. La somme des valeurs par ligne correspond au nombre d'échantillons de test par classe. Le taux de classification par classe est donné par la valeur à la diagonale divisée par la somme des valeurs par ligne.

**Cas de deux classes :** L'évaluation d'un système de classification binaire se base sur le calcul du taux de classification, de la sensibilité et de la spécificité.

Le taux de classification est calculé par l'équation :

$$T = \frac{N1}{N} \quad (1.3)$$

Où  $T$  désigne le taux de classification,  $N1$  est le nombre d'échantillons bien classifiés et  $N$  le nombre total des échantillons.

En utilisant cette relation, nous pouvons déduire le taux d'erreur,  $TE$ , par la relation suivante :

$$TE = 1 - T \quad (1.4)$$

Dans le cas de deux classes  $C1$  et  $C2$ , la matrice de confusion est illustrée dans le tableau suivant :

Classe réelle	Classe prédite	
	C1	C2
C1	VP	FN
C2	FP	VN

**Tableau 1: Matrice de confusion dans le cas binaire**

Avec :

- VP (Vrai positif) : Éléments de la classe C1 correctement classifiés;
- VN (Vrai négatif) : Élément de la classe C2 correctement classifié;
- FP (Faux positif) : Éléments de la classe C2 classifiés comme étant de la classe C1;
- FN (Faux négatif) : Éléments de la classe C1 classifiée comme étant de la classe C2.

Cette matrice permet de déduire les paramètres suivants pour l'évaluation d'un classificateur binaire:

$$\text{Précision} = \frac{(\text{VP} + \text{VN})}{N}$$

$$\text{Sensibilité} = \frac{\text{VP}}{(\text{VP} + \text{FN})}$$

$$\text{Spécificité} = \frac{\text{VN}}{(\text{VN} + \text{FP})}$$

## 2.3. Techniques de classification

Nous distinguons deux catégories de techniques de classification : les techniques supervisées et les techniques non supervisées.

### La classification supervisée

Dans le cas de classification supervisée (en anglais *supervised learning*), le système observe des couples de types entrée-sortie et apprend une fonction qui permet d'aboutir à la sortie à partir de l'entrée. Cette phase est appelée phase d'apprentissage ou d'entraînement. Les données de sortie correspondantes sont appelées classes.

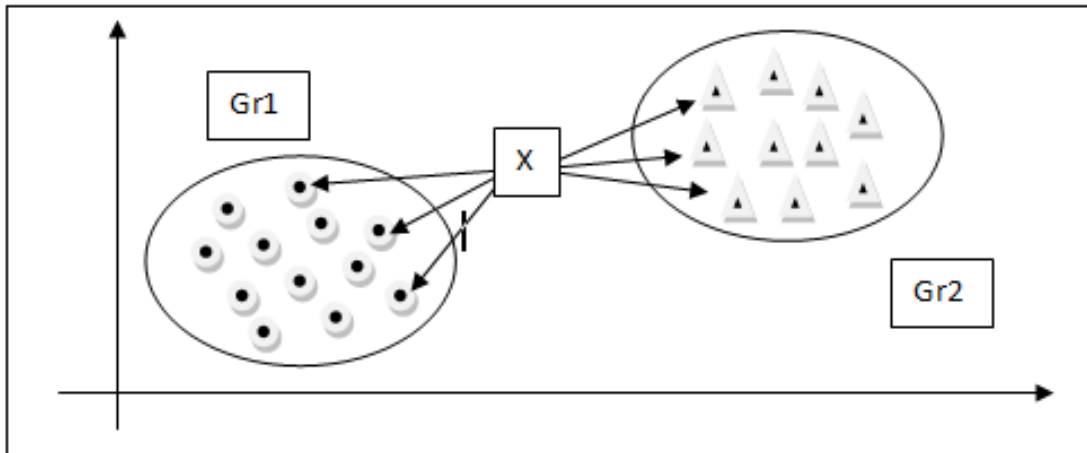
Il existe différents algorithmes pour la classification supervisée. Parmi ces algorithmes on cite, le K plus proche voisin (*KNN : K-Nearest Neighbor*), les arbres de décision et les machines à vecteur de support (SVM).

- **K plus proche (KNN)**

Le principe de l'algorithme KNN est de classer les exemples dont la classe est inconnue en se basant sur leur similarité avec un ensemble d'échantillons de référence. La similarité est déterminée par un calcul de distance. La distance euclidienne décrite dans l'équation suivante est celle la plus utilisée :

$$d(X, u) = \sqrt{\sum_{i=1}^n (x_i - u_i)^2} \quad (1.2)$$

Où  $d$  est la distance euclidienne,  $X = (x_1, x_2, \dots, x_n)$  est l'élément à classer et  $u$  est un élément de la base de référence.  $n$  étant la dimension du vecteur de caractéristiques.



**Figure 7 :** Distance euclidienne entre X et les éléments des deux classes Gr1 et Gr2

La figure 7 illustre le principe du KNN; pour déterminer la classe d'un échantillon X à l'aide du classificateur KNN. Nous cherchons les plus proches voisins par le calcul de la distance euclidienne entre X et tous les éléments des deux classes Gr1 et Gr2. L'élément X est affecté à la classe où le nombre des voisins trouvés est majoritaire.

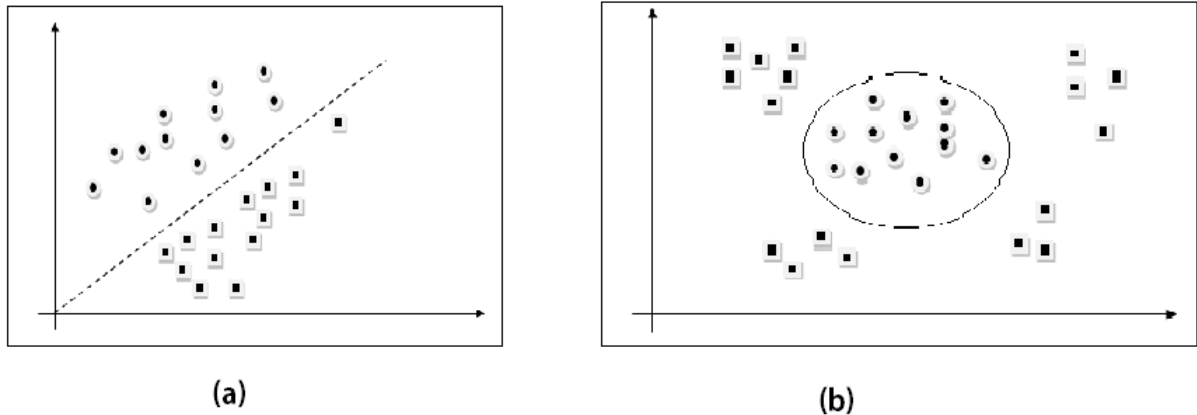
L'algorithme KNN présente plusieurs avantages tels que la facilité de son intégration et de sa mise en œuvre et sa robustesse. Cependant, il a des inconvénients tels que le temps de classification qui est important puisqu'il faut évaluer la distance entre l'élément à classer et les éléments de références de toutes les classes à chaque fois. L'algorithme KNN est, de plus, sensible aux attributs non pertinents et corrélés.

- **Les arbres de décision**

L'apprentissage automatique par arbres de décisions permet de représenter graphiquement le système de classification de façon à rendre la compréhension de son fonctionnement plus facile pour l'utilisateur. Les arbres de décision permettent de classer un ensemble d'individus en se basant sur la pertinence des attributs par rapport à leurs classes. Les arbres de décision présentent des avantages, à savoir, la facilité de mise en œuvre, la facilité de prise de décision et la simplification des décisions complexes.

- **Les machines à vecteurs de support (SVM)**

La classification basée sur les machines à vecteurs de support est relativement récente ; elle a été créée par Vapnik Vladimir en 1995. Le principe de base des SVM consiste à maximiser la distance entre deux classes pour résoudre le problème de discrimination avec un séparateur linéaire ou non linéaire [22].



**Figure 8 : Discrimination à deux classes**  
**(a) : Séparateur linéaire - (b) : Séparateur non linéaire**

### **La classification non supervisée**

Contrairement à la classification supervisée, la classification non supervisée se base sur des algorithmes qui permettent de classifier en regroupant les individus similaires entre eux sans passer par la phase d'apprentissage. Parmi ces algorithmes on trouve le K moyennes (*K-means*) et celui de carte Kohonen (réseaux de neurones).

- **K-moyennes**

Créé par Queen en 1997, l'algorithme K-moyennes permet la minimisation alternée en regroupant en K classes un ensemble d'individus sachant que K est fixé par l'utilisateur. Le choix de K n'est pas universel, plus on augmente la valeur de K plus la distorsion diminue [23].



- **Carte de Kohonen**

Il s'agit d'un algorithme original défini par Teuvo Kohonen aux années 80 se basant sur la notion de voisinage entre les classes et l'appartenance des observations voisines, dans une dimension  $P$ , dans une même classe ou dans des classes voisines [24].

## **2. 4. Classification de l'activité physique**

Plusieurs études ont appréhendé la classification automatique de l'activité physique. La différence entre ces méthodes se situe au niveau des capteurs utilisés, des caractéristiques extraites et des techniques de classification utilisées. Ce tableau de synthèse qui dresse l'état de l'art des systèmes de classification de l'activité physique se base sur l'étude de Robertas Damasevicius et al [25] dans laquelle il a résumé les différentes méthodes de classification ainsi que les caractéristiques utilisées.

D'après le tableau 2, nous constatons que, la majorité des études se sont basées sur l'accéléromètre afin de se servir des valeurs des trois composantes d'accélérations et/ou de l'accélération linéaire. Dans ces études, les taux de classification obtenus varient entre 64,5% et 99,1% selon le nombre d'activités physiques à classer et le type de classificateur. Nous remarquons également une variété de caractéristiques extraites que ce soit dans le domaine temporel ou le domaine fréquentiel. Les meilleures performances en termes de taux de bonne classification obtenu est celui de l'étude de Mannini et Sabatini [38]. Dans cette étude, les données d'accélération ont été utilisées pour classer 21 activités physiques en utilisant le modèle de Markov après l'extraction des caractéristiques basée sur l'énergie, l'entropie et les coefficients de corrélation. D'autres études ont permis de classer, avec un taux de classification important, les activités physiques en se basant sur les données de l'accéléromètre combinées avec d'autres capteurs. Il est à noter que les capteurs utilisés dans la majorité de ces études sont des capteurs adhésifs installés en contact direct avec les participants pour une meilleure qualité de collecte des données. Cet aspect est contraignant pour les utilisateurs et rend l'applicabilité des méthodes développées contraignante. Ainsi, nous proposons d'investiguer des capteurs intégrés dans des téléphones intelligents installés dans la poche de l'utilisateur ou dans une ceinture dédiée aussi fréquemment utilisée par les utilisateurs.

Auteur	Activités	Données des capteurs	Caractéristiques	Sélection de caractéristiques	Méthode de classification	Taux
Atallah et al. [26]	S'allonger, cuisiner, manger et boire, socialiser, lecture, s'habiller, marcher sur tapis roulant, marcher, passer l'aspirateur, essuyer les tables, courir, cyclisme, s'asseoir et se lever, se coucher et se lever.	Accélération	Entropie moyenne sur 3 axes, fréquence principale FFT (En moyenne) sur 3 axes, énergie en 0.2Hz fenêtre centrée autour de la fréquence principale sur totale énergie FFT (moyenne 3-axes), moyenne de covariance croisée entre tous les 2 axes	Relief, Simba et MRMR	KNN, classificateur Baiesien	90%
Bayat et al. [27]	Courir, marche lente, marche rapide, danse aérobique, monter les escaliers, descendre les escaliers	Triaxial Accéléromètre	Moyenne le long de l'axe z, MinMax, STD, et RMS Am, APF le long de l'axe x, l'axe y et l'axe z, VarAPF, STD le long de l'axe x, l'axe y, et l'axe des z, le long de RMS axe des x, l'axe y et l'axe z, corrélation entre l'axe z et l'axe y et min./max le long de l'axe x, l'axe y, et l'axe z	Regroupement des caractéristiques	Multilayer perceptron, SVM, Random Forest, and Logit Boost	81%–91%
Berchtold et al. [28]	Debout, assis, couché, marcher, monter les escaliers, le cyclisme, et étant stationnaire	Accéléromètre	Variance, moyenne	Aucune	Fuzzy inference	97.3%
Capela et al. [29]	Assis, debout, s'allonger, ramper vers le haut et le bas, monter et descendre les escaliers, transition entre les activités	Accélération linéaire, gravité, capteur de vitesse	Moyenne, écart-type, aplatissement, mouvement moyenne, matrice de covariance, dissymétrie, taux zéro croix, et le taux moyen croix	Aucune	Naive-Bayes, SVM, arbre de decision j48	97%
Gupta and	Sauter, courir, marcher, assis,	Triaxial	L'énergie, l'entropie, moyenne,	Relief, SFFS	kNN, Naive	98%

Dallas [30]	assis à debout, et debout à genoux	accéléromètre	variance, X-Z-énergie et différence max de l'accélération		Bayes	
Henpraserttae et al. [31]	Assis, coucher, debout, marcher	Accéléromètre	Moyenne, Écart-type	Aucune	Rules and threshold based classification	90%
Hoque and Stankovic [32]	Quitter la maison, utiliser les toilettes, prendre une douche, dormir, la préparation du petit-déjeuner, préparation du dîner, faire collation, boire, utiliser la machine à laver, utiliser la lave-vaisselle	Location sensors (open/closed)	Magnétomètre	Aucune	Custom clustering method	64.5%–89.9%
Iso and Yamazaki [33]	Marcher, courir, monter/descendre les escaliers, marche rapide	Accéléromètre	Composants ondelettes, Périodogrammes et entropie	Aucune	Probabilités bayésiennes	80%
Kose et al. [34]	Marcher, courir, vélo, assis, debout	Accéléromètre	Min., Max., Moyenne, variance, coefficients FFT, et autocorrélation	Aucune	Clustered kNN	95.2%–97.5%
Kwapisz et al [35]	Marche, jogging, monter/descendre les escaliers, assis, debout	Accéléromètre	Moyenne, std. dev., moyenne différence absolue, moyenne accélération résultante, le temps entre les piques et binned distribution	Aucune	Arbre de décision, logistic regression, and MNN	91.7%
Lane et al. [36]	Conduire, stationner, courir, marcher	GPS, Accéléromètre, et microphone	Moyenne, variance	Aucune	Naive-Bayes	85–98%
Lee and Cho [37]	Être debout, marcher, courir, monter/descendre les escaliers, shopping, et prendre le bus	Accéléromètre	valeurs accélération pour les axes x, y et z.	Aucune	Hierarchical HMM	70%–90%

Mannini and Sabatini [38]	Marcher, marcher en transportant des objets, assis & détente, travailler sur ordinateur, debout, manger ou boire, regarder la télévision, la lecture, la course, le vélo, étirage, la force la formation, le lavage, passer l'aspirateur, le pliage blanchisserie, couché et détente, se brosser les dents, monter les escaliers, prendre ascenseur, prendre escalier mécanique	Accélération	composante continue, l'énergie, domaine fréquentiel entropie, et les coefficients de corrélation	SFFS (Pudil algorithm)	Continuous emissions, Hidden Markov Model	99.1%
Mathie et al. [39]	les différents mouvements humains y compris repos, marcher et chutter	Triaxial Accélération sensor	zone intégrée sous la courbe	Aucune	Arbre de décision binaire	97.7% (sensitivity) 98.7% (specificity)
Maurer et al. [40]	Marcher, debout, assis, courir, monter/descendre les escaliers	Multiple sensors	Moyenne, moyenne quadratique, écart-type, variance, moyenne absolue écart, cumulative histogramme, percentiles, l'interquartile, ZCR, MCR	Correlation based Feature Selection (CFS)	Arbre de décision (C4.5 algorithm), <i>k</i> -Nearest Neighbor, Naive-Bayes	80%–92%
Miluzzo et al. [41]	Assis, debout, marcher, courir	Accéléromètre, GPS, audio	DFT, les caractéristiques de la FFT, moyenne, std et le nombre de pics par unité	Aucune	Arbre de décision	79%

Parkka et al. [42]	Allongé, vélo, assis, debout, courir, marcher	GPS, audio, altitude, EKG, accelerometer, compass, humidity, light, temperature, heart rate, pulse, respiratory effort, and skin resistance	Fréquence de pics de haut en bas de l'accélération, la médiane de haut en bas de l'accélération, puissance de pics de l'accélération, la variance l'accélération, la somme de variances de poignet 3D l'accélération et la puissance rapport des bandes de fréquences 1-1.5Hz et 0.2-5Hz mesurée à partir du magnétomètre	Heuristic	Arbre de décision	86%
Saponas et al. [43]	Marcher, jogging	Accéléromètre	124 caractéristiques: Nike + iPod Payload Packet, l'ampleur (Moyenne, std. Dev., Min., max., et min. moins max.), la fréquence (en énergie de chacune des 10 premières composantes de fréquence DFT, l'énergie de chaque bande de 10 fréquences composants, la plus grande composante de fréquence, et indice de la plus grande composante de fréquence)	Aucune	Naive-Bayesian Network	97.4% (within person), 99.48% (cross-person)
Siirtola and Roning [44]	Marcher, courir, vélo, conduire, assis, et debout	Accelerometer	Magnitude, std., moyenne, min., max., percentiles (10, 25, 50, 75 et 90), et la somme et somme des carrés de observations dessus / dessous percentile (5, 10, 25, 75, 90, et 95) d'une ampleur accélération et la somme carrée de X et Z	Aucune	Arbre de décision + kNN/QDA	95%

Sohn et al. [45]	Marcher, conduire	GPS	Corrélation de rang, variance, moyenne, La distance euclidienne au cours d'une Fenêtre de mesures	Aucune	Logistic regression	85%
Yang [46]	Assis, debout, marcher, courir, conduire, vélo	Accelerometer	Moyenne, std., ZCR, 75e percentile, interquartile, spectre barycentre, l'entropie, et corrélation croisée	Aucune	Arbre de décision, Naive-Bayes, kNN, and SVM	90%
Zhu and Sheng [47]	Assis, debout, coucher, marcher, assis à debout, debout à assis, coucher à assis, et assis à coucher	3D acceleration	Moyenne, variance	Aucune	Neural network ensemble	67%–98%

**Tableau 2 : Résumé des travaux effectués dans le domaine de classification des activités physiques**  
**Tiré de : Robertas Damasevicius et al [25]**

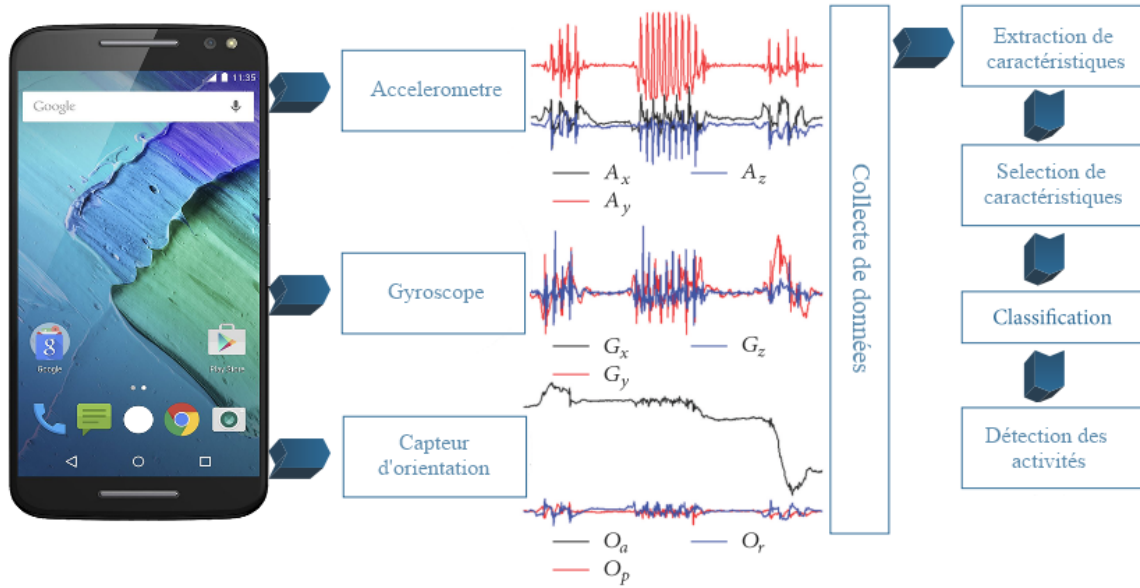
# Chapitre3

## Systeme de classification de l'activité physique

Ce chapitre décrit le système de classification de l'activité physique que nous proposons. Nous décrirons dans un premier temps la méthodologie adaptée (Section 3.1) ensuite nous présentons les résultats expérimentaux (section 3.2).

### 3.1. Méthodologie

Tel que décrit précédemment, l'objectif principal de ce mémoire est de développer un système de classification automatique des activités physiques. Pour répondre à cet objectif, nous avons adopté la méthodologie suivante (décrite dans la figure 9): Dans un premier temps, nous avons développé deux applications une application mobile et application web. Ces applications permettent de collecter et de gérer les données provenant des téléphones intelligents. Une fois les données collectées nous avons procédé à l'extraction des caractéristiques et la conception du système de classification en vue de la détection de l'activité physique. Nous avons testé trois classificateurs, dont les K-plus proches voisins, le classificateur bayésien, les arbres de décision et la machine à vecteur de support (SVM). La validation s'est basée sur la matrice de confusion et les taux de bonne classification.



**Figure 9 : Diagramme en bloc de la méthodologie adoptée**

### **A. Conception de la base de données**

La première étape consistait à concevoir et installer un système de gestion de base de données MySQL sur un serveur distant où toutes les données sont enregistrées.

La conception de la base de données a été basée sur les besoins de la recherche, à savoir, la création des profils des participants et l'enregistrement des données des différents capteurs (téléphones intelligents et héxoskin). La méthode MERISE a été appliquée lors de la conception de la base de données. La figure 10 illustre le modèle physique de la base de données et les liaisons entre les différentes tables.



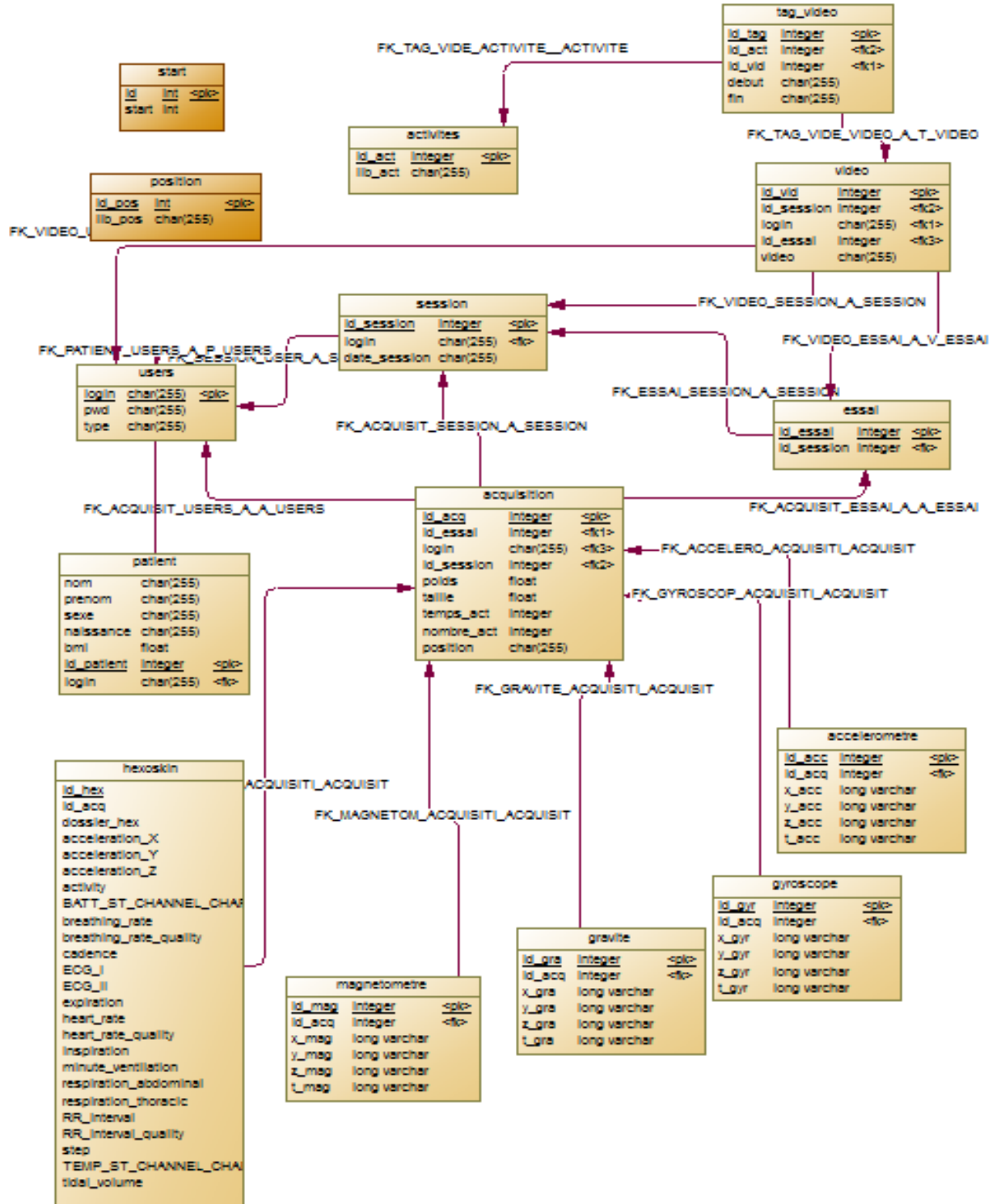


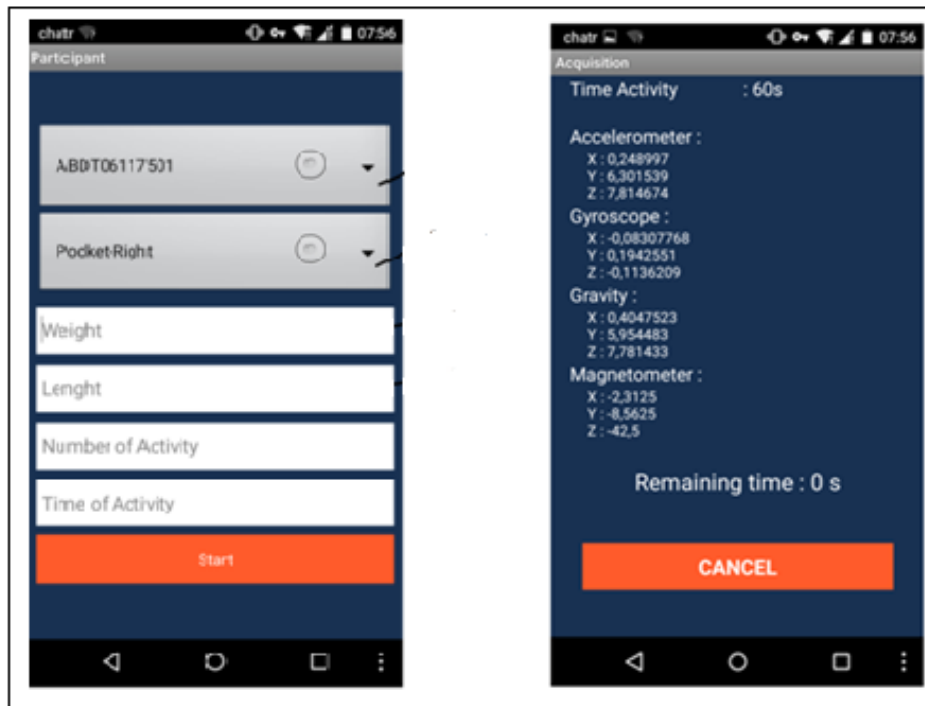
Figure 10: Modèle physique des données

## B. Développement d'interfaces

L'étape suivante était le développement des deux interfaces : La première est une application mobile qui a aussi servi d'interface de collecte de données. La deuxième est une application web pour permettre l'interaction entre les pages web et la base de données.

- **Application mobile**

Cette application a été développée en utilisant le langage *C#* comme langage de programmation avec le logiciel Xamarin qui offre un SDK répondant à l'environnement du développement Android. Aussi le protocole JSON a été utilisé pour la communication avec le serveur des données (MySQL). L'application mobile a permis suite à l'entrée des informations concernant un participant (à savoir, l'âge, le poids et la taille) de collecter les données des quatre capteurs (accéléromètre, gyroscope, gravité et magnétomètre).



**Figure 11 : Capture d'écran pour l'application mobile**

- **Application web**

La console d'administration web développée a permis, d'une part, de créer le profil d'un nouveau participant dans la base de données et, d'autre part, de visualiser les données provenant des quatre capteurs (accéléromètre, gyroscope, gravité et magnétomètre) sous la forme de graphiques.

Elle se compose des cinq onglets suivants:

- Participant : Pour ajouter, modifier ou supprimer un participant
- Données : Pour visualiser les données sous forme de graphique
- Utilisateurs : Pour ajouter, modifier ou supprimer un utilisateur de la console d'administration, avec possibilité d'accorder des droits d'accès pour chaque onglet
- Setting : pour permettre à l'administrateur de faire une sauvegarde, en ligne, de la base de données, d'ajouter ou supprimer des activités et des positions pour les téléphones intelligents, d'ajouter ou supprimer des plages d'adresses IP qui auront la possibilité de se connecter à la console d'administration et de visualiser l'historique de connexion des différents utilisateurs.
- Se déconnecter : pour se déconnecter de la console d'administration

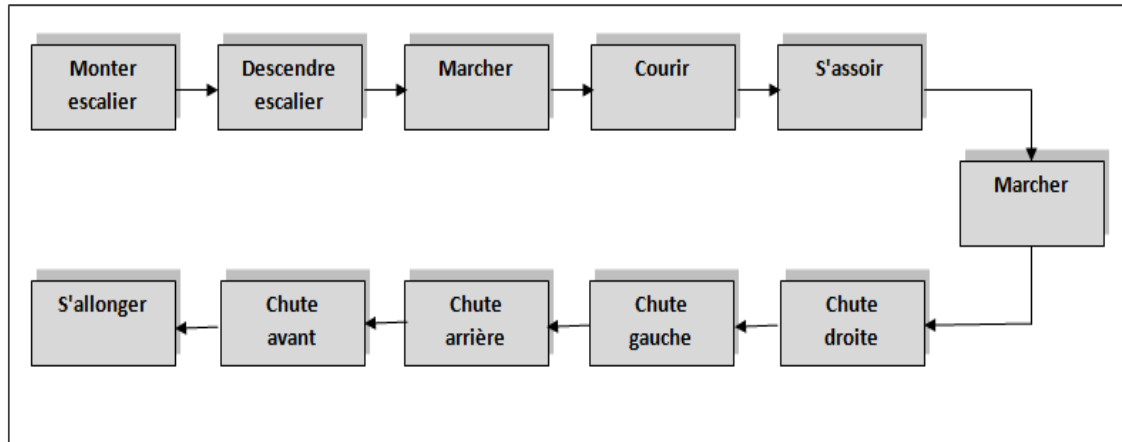
L'accès aux différents onglets est limité par un droit d'accès, que seul l'administrateur a la possibilité d'attribuer à un utilisateur.

### **C. Collecte de données :**

Après avoir enregistré les informations du participant (Nom, Prénom, date de naissance et sexe) dans la base de données via l'application web, un code est généré automatiquement et remplace le nom et le prénom du participant. Ce code est utilisé lors du lancement de l'application Android qui est installée dans les deux téléphones intelligents.

La collecte de données a eu lieu dans les locaux du CRCHUM (Centre de Recherche du Centre Hospitalier de l'Université de Montréal) avec la participation de sept personnes. Chaque participant est appelé à faire, cinq fois, un cycle contenant dans l'ordre les 7

activités physiques suivantes (voir figure 12) : monter (C11) et descendre l'escalier (C12), marcher (C13), courir (C14), s'asseoir (C15), chutes (C16) et s'allonger (C17). Les classes C16 regroupe quatre types de chutes : gauche, droite, arrière et avant.

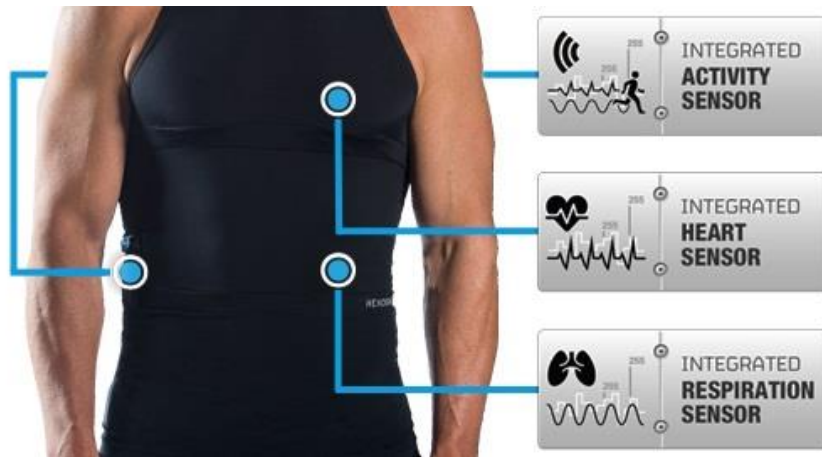


**Figure 12 : Cycle des activités physiques**

Durant le cycle d'activité, les téléphones intelligents enregistrent les données des capteurs (accéléromètre, gyroscope) et les sauvegardent dans une base de données. Ces données sont exploitées via l'application web qui les affichera sous forme de graphe et génère un fichier script pour être utilisée directement sur le logiciel de traitement de données MATLAB.

En plus des téléphones intelligents, dont un est placé dans la poche du participant et l'autre sous le dos (voir figure 14), le participant est habillé du textile intelligent (Héxoskin) pour permettre d'enregistrer les données des capteurs installés dans l'Héxoskin (Figure 13). Ces données n'ont pas été utilisées dans le cadre de ce projet; elles font actuellement l'objet d'un autre projet dont l'objectif est la classification de l'activité en utilisant les signaux physiologiques. Nous en parlons, ici, uniquement parce qu'ils ont fait partie du protocole de collecte des données.

À la fin de chaque cycle, l'application enregistre les données dans la base de données installée sur un serveur distant. À noter que les applications Android installées dans les deux téléphones doivent être lancées et arrêtées en même temps via une commande externe permettant le lancement et l'arrêt de l'acquisition.



**Figure 13 : Données collectées par l'Héxoskin**



**Figure 14 : Position des téléphones intelligents lors de la collecte des données**

- **Données collectées**

Les données collectées sont les suivantes :

- Les données anthropométriques des participants (poids, grandeur, sexe).
- Les données accélérométriques des participants enregistrées à partir des capteurs des deux terminaux mobiles.
- Les données de l'activité cardiaque (fréquence cardiaque, HRV (permettant d'estimer le stress et la fatigue), récupération de la fréquence cardiaque, etc.) mesurées par l'hexoskin.

- Les données de l'activité respiratoires (taux de respiration (RPM) et la ventilation minute (L/min)), également, mesurées par l'hexoskin.
- Des caméras vidéos permettent l'enregistrement des séquences vidéos afin d'étiqueter les différentes classes d'activités physiques (marcher, debout, courir, dormir, chute, monter et descendre les escaliers). Ces étiquettes ont servi de références pour la validation des méthodes d'analyse et de classification de données.

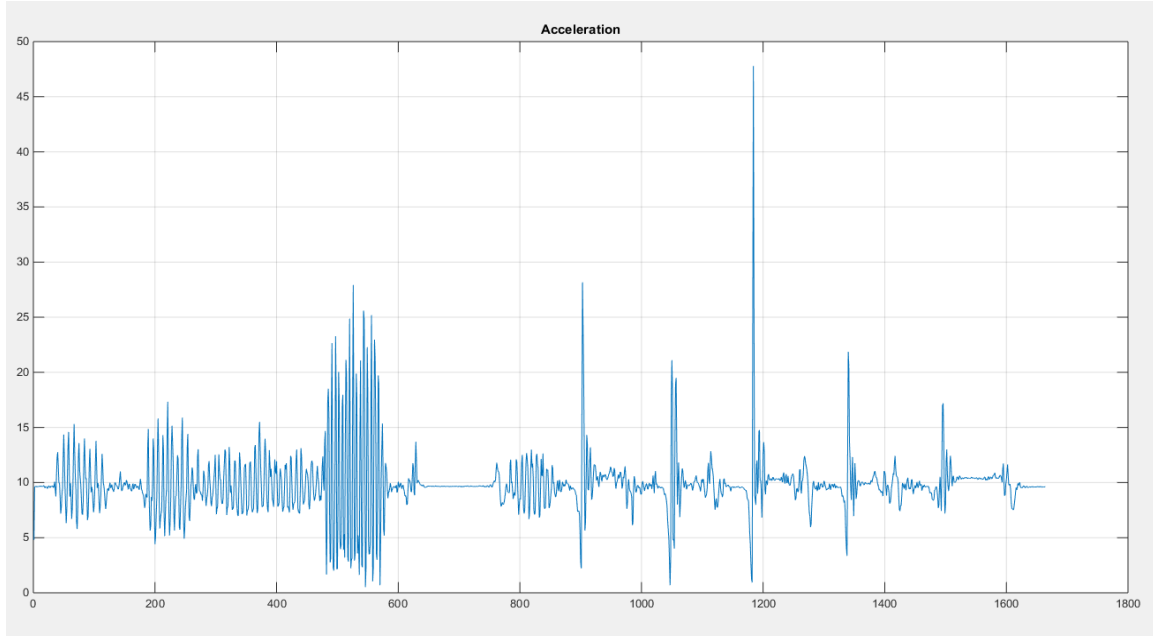
- **Considération éthique**

Toute information recueillie au cours de cette étude a été traitée de manière confidentielle. Les données auxquelles nous avons eu accès sont anonymisées. Ce projet a été soumis aux comités d'éthique de la recherche du Centre hospitalier de l'Université de Montréal, de l'École de technologie supérieure et de la TELUQ. Les principes éthiques fondamentaux ainsi que les lignes directrices des conseils ont été respectés durant ce projet. Tous les participants recrutés durant cette étude participent de façon volontaire et pouvaient se retirer à tout moment. La nature des évaluations, leur déroulement ainsi que les risques relatifs à la participation à cette étude ont été expliqués aux participants lors de la signature du formulaire de consentement.

## **D. Système de classification**

- **Prétraitement :**

Comme mentionné dans le chapitre 1, le processus de classification passe en premier temps par la phase de prétraitement. Dans cette étude, on a opté pour le filtre de *Kalman* afin de supprimer le bruit d'acquisition. La figure 15 illustre l'amplitude du signal obtenu d'accélération après filtrage.



**Figure 15 : Signal d'accélération linéaire filtré pour un participant  
(position : Ceinture)**

- **Extraction des caractéristiques :**

Les signaux exploités sont ceux provenant de l'accéléromètre et du gyroscope. Soit  $S_n$  un signal de l'ensemble des signaux  $\{A_x, A_y, A_z, A, G_x, G_y, G_z, G\}$  avec  $n$  la dimension du signal considéré.  $A_x, A_y$  et  $A_z$  désignent les composantes tridimensionnelles de l'accélération et  $G_x, G_y$  et  $G_z$  les composantes tridimensionnelles du gyroscope.  $A$  et  $G$  désignent la norme de l'accélération et du gyroscope. Les caractéristiques ont été extraites sur chacune des composantes. Nous avons considéré l'ensemble des caractéristiques suivantes : la valeur moyenne du signal, l'amplitude, le minimum, le maximum, l'étendue ainsi que le facteur d'asymétrie. Ces caractéristiques sont données par les équations suivantes :

$$\text{maximum}(s) = \max_n(S_n)$$

$$\text{minimum}(s) = \min_n(S_n)$$

$$mean(s) = \mu = \frac{1}{N} \sum_{n=1}^N S_n$$

$$Range(s) = \max_n(S_n) - \min_n(S_n)$$

$$asym\acute{e}trie(s) = \frac{1}{N\sigma^3} \sum_{n=1}^N (S_n - \mu)^3$$

Avec  $variance(s) = \sigma^2 = \frac{1}{N} \sum_{n=1}^N (S_n - \mu)^2$

Les caractéristiques extraites sont utilisées par la suite pour des fins de classifications.

- **Classification :**

Dans cette étude, trois modèles de classification ont été utilisés, à savoir, le KNN (K plus voisin proche), le classificateur Bayésien et les arbres de décision. Les classificateurs ont été développés pour les données collectées depuis les deux appareils mobiles (installés dans la poche et dans la ceinture en bas du dos).

### 3.5. Résultats expérimentaux : Tests et validation

La base de données contient 350 échantillons. En effet, chaque participant est appelé à faire, 5 fois, un cycle contenant dans l'ordre les activités physiques suivantes : monter (Cl1) et descendre l'escalier (Cl2), marcher (Cl3), courir (Cl4), s'asseoir (Cl5), chutes (Cl6) et s'allonger (Cl7). Les classes Cl6 regroupent quatre types de chutes : gauche, droite, arrière et avant. Nous avons ainsi obtenu 35 échantillons par classes (7 participants fois 5 répétitions) sauf pour la classe chutes, pour laquelle nous avons obtenu 140 échantillons (7 participants fois 5 répétitions fois 4 types de chutes)

Les systèmes de classification développés ont été évalués en termes de taux de bonne classification. Les vecteurs de caractéristiques sont formés par la concaténation de l'ensemble des caractéristiques précédemment décrites (la valeur moyenne du signal, l'amplitude, le minimum, le maximum et l'étendue ainsi que l'asymétrie). Les



caractéristiques discriminantes ont été choisies de manière expérimentale parmi l'ensemble de toutes les caractéristiques.

Les tests ont été faits selon le processus de Leave-one-out cross validation pour chacun des deux emplacements des téléphones intelligents (poche et bas de dos (ceinture)). Le tableau 3 résume les résultats obtenus pour les trois classificateurs développés. On note que les taux de classification sont toujours plus élevés lorsque le téléphone intelligent est placé au niveau de la ceinture.

Classifieur	Taux de classification en %	
	Poche	Ceinture
KNN	80,29%	88%
Arbre de décision	78,57%	79,71%
Classificateur Bayésien	61,43%	67,43%

**Tableau 3 : Taux de classification des différents classificateurs dans les deux positions (poche et ceinture)**

Ce résultat est prévisible puisque dans cet emplacement le téléphone est plus stable. Le taux de classification s'élève à 88% en utilisant le KNN. Ce résultat est très encourageant et se compare favorablement à la littérature. La comparaison entre les trois méthodes de classification démontre que le KNN donne les meilleures performances. De plus, ce classificateur est très simple à implémenter dans un contexte d'application mobile sur des terminaux mobiles.

Pour une meilleure investigation des résultats, nous avons représenté les matrices de confusion des trois classificateurs dans chacun des deux emplacements (tableau 4 à 9). La dernière colonne de chaque tableau correspond au taux de classification par classe.

		Classes prédites							
		CI1	CI2	CI3	CI4	CI5	CI6	CI7	
Classes Réelles	CI1	30	3	0	0	1	1	0	85,71%
	CI2	4	30	0	1	0	0	0	85,71%
	CI3	1	0	34	0	0	0	0	97,14%
	CI4	1	2	0	29	2	0	1	82,86%
	CI5	2	1	1	4	26	0	1	74,29%
	CI6	4	0	0	0	3	129	4	92,14%
	CI7	0	0	0	0	2	3	30	85,71%

**Tableau 4 : Matrice de confusion du classificateur KNN à partir de données collectées au niveau de la ceinture. CI1 : monter les escaliers, CI2 : descendre les escaliers, CI3 : marcher, CI4 : courir, CI5 : s'asseoir, CI6 : chutes et CI7 : s'allonger.**

Les classes CL6 regroupent 4 types de chutes : gauche, droite, arrière et avant.

La colonne de droite correspond aux pourcentages de classification correcte de chaque classe

		Classes prédites							
		CI1	CI2	CI3	CI4	CI5	CI6	CI7	
Classes Réelles	CI1	28	7	0	0	0	0	0	80,00%
	CI2	8	22	1	0	4	0	0	62,85%
	CI3	0	1	34	0	0	0	0	97,14%
	CI4	1	0	0	30	2	0	2	85,71%
	CI5	2	4	0	3	19	6	1	54,28%
	CI6	0	0	0	1	9	122	8	87,14%
	CI7	0	0	0	1	1	7	26	74,28%

**Tableau 5 : Matrice de confusion du classificateur KNN à partir de données collectées au niveau de la poche. CI1 : monter les escaliers, CI2 : descendre les escaliers, CI3 : marcher, CI4 : courir, CI5 : s'asseoir, CI6 : chutes et CI7 : s'allonger.**

Les classes CL6 regroupent 4 types de chutes : gauche, droite, arrière et avant.

La colonne de droite correspond aux pourcentages de classification correcte de chaque classe

		Classes prédites							
		CI1	CI2	CI3	CI4	CI5	CI6	CI7	
Classes Réelles	CI1	28	4	0	0	3	0	0	84,84%
	CI2	9	25	0	0	1	0	0	75,75%
	CI3	4	0	30	0	1	0	0	90,91%
	CI4	0	2	1	24	2	0	6	72,72%
	CI5	7	4	1	2	11	0	10	33,33%
	CI6	30	4	0	1	2	68	35	51,15%
	CI7	0	0	0	0	2	4	29	87,87%

**Tableau 6 : Matrice de confusion du classificateur Bayésien à partir de données collectées au niveau de la ceinture. CI1 : monter les escaliers, CI2 : descendre les escaliers, CI3 : marcher, CI4 : courir, CI5 : s'asseoir, CI6 : chutes et CI7 : s'allonger.**  
**Les classes CL6 regroupent 4 types de chutes : gauche, droite, arrière et avant.**  
**La colonne de droite correspond aux pourcentages de classification correcte de chaque classe**

		Classes prédites							
		CI1	CI2	CI3	CI4	CI5	CI6	CI7	
Classes Réelles	CI1	29	5	0	0	1	0	0	87,87%
	CI2	5	29	0	1	0	0	0	87,87%
	CI3	0	1	31	3	0	0	0	93,93%
	CI4	1	0	0	28	2	0	4	84,84%
	CI5	1	5	0	6	8	1	14	24,24%
	CI6	0	1	0	0	5	83	51	62,87%
	CI7	0	1	0	0	0	6	28	84,84%

**Tableau 7 : Matrice de confusion du classificateur Bayésien à partir de données collectées au niveau de la poche. CI1 : monter les escaliers, CI2 : descendre les escaliers, CI3 : marcher, CI4 : courir, CI5 : s'asseoir, CI6 : chutes et CI7 : s'allonger.**  
**Les classes CL6 regroupent 4 types de chutes : gauche, droite, arrière et avant.**  
**La colonne de droite correspond aux pourcentages de classification correcte de chaque classe**

		Classes prédites							
		CI1	CI2	CI3	CI4	CI5	CI6	CI7	
Classes Réelles	CI1	25	5	0	0	5	0	0	75,75%
	CI2	5	28	0	0	2	0	0	84,84%
	CI3	0	0	31	3	1	0	0	93,93%
	CI4	3	1	1	22	5	2	1	66,66%
	CI5	4	4	0	2	22	3	0	66,66%
	CI6	1	1	0	1	3	123	11	93,18%
	CI7	0	0	0	1	2	8	24	72,72%

**Tableau 8 : Matrice de confusion du classificateur utilisant les arbres de décision à partir de données collectées au niveau de la ceinture. CI1 : monter les escaliers, CI2 : descendre les escaliers, CI3 : marcher, CI4 : courir, CI5 : s'asseoir, CI6 : chutes et CI7 : s'allonger.**

**Les classes CL6 regroupent 4 types de chutes : gauche, droite, arrière et avant.**

**La colonne de droite correspond aux pourcentages de classification correcte de chaque classe**

		Classes prédites							
		CI1	CI2	CI3	CI4	CI5	CI6	CI7	
Classes Réelles	CI1	26	7	0	1	1	0	0	78,78%
	CI2	5	24	0	0	6	0	0	72,72%
	CI3	0	0	31	0	4	0	0	93,93%
	CI4	1	0	0	28	3	0	3	84,84%
	CI5	0	4	4	0	20	2	5	60,60%
	CI6	0	0	0	1	3	126	10	95,45%
	CI7	0	1	0	1	1	8	24	72,72%

**Tableau 9 : Matrice de confusion du classificateur utilisant les arbres de décision à partir de données collectées au niveau de la poche. CI1 : monter les escaliers, CI2 : descendre les escaliers, CI3 : marcher, CI4 : courir, CI5 : s'asseoir, CI6 : chutes et CI7 : s'allonger.**

**Les classes CL6 regroupent 4 types de chutes : gauche, droite, arrière et avant.**

**La colonne de droite correspond aux pourcentages de classification correcte de chaque classe**

Le tableau 4 illustre la matrice de confusion du classificateur KNN lorsque le terminal mobile est installé au niveau de la poche pour lequel nous avons obtenu un taux de bonne classification de 88%. Les étiquettes de ligne désignent les classes réelles et les étiquettes en colonne désignent les étiquettes prédites. L'analyse de cette matrice montre que la classe Cl5 (s'asseoir) qui est confondue à 17,14% (6/35) avec la classe Cl6 (chute). Cette confusion est due à la nature des caractéristiques qui se basent entre autres sur l'amplitude de l'accélération qui se trouve à être sensiblement proches entre ces deux classes. Pour les autres classes, les taux de classification par classe sont assez élevés ce qui confirme que l'utilisation des capteurs d'un terminal mobile permettent de classer les activités physique d'un utilisateur avec un taux de réussite assez élevé.

# CONCLUSION

Cette étude a permis de développer un système de classification des activités physiques en utilisant deux appareils mobiles dans deux positions différentes (dans la poche et attachée à une ceinture au dos). Pour ce faire, nous avons conçu un système de gestion de la base des données collectées et nous avons développé deux interfaces : La première est une application mobile qui a aussi servi d'interface de collecte de données. La deuxième est une application web pour permettre l'interaction entre les pages web et la base de données.

La collecte de données a eu lieu dans les locaux du CRCHUM (Centre de Recherche du Centre Hospitalier de l'Université de Montréal) avec la participation de sept personnes. Chaque participant est appelé à faire, cinq fois, un cycle contenant dans l'ordre les 7 activités physiques suivantes : monter (C11) et descendre l'escalier (C12), marcher (C13), courir (C14), s'asseoir (C15), chutes (C16) et s'allonger (C17). Les classes C16 regroupe quatre types de chutes : gauche, droite, arrière et avant. Durant le cycle d'activité, les téléphones intelligents enregistrent les données des capteurs (accéléromètre, gyroscope) et les sauvegardent dans une base de données.

Nous avons par la suite développé trois modèles de classification ont été utilisés, à savoir, le KNN (K plus voisin proche), le classificateur Bayésien et les arbres de décision. Les systèmes de classification développés ont été évalués en termes de taux de bonne classification.

Les résultats obtenus montrent que la position de la poche est plus intéressante que celle de la ceinture vu que les taux de classification obtenu est plus élevé. Les résultats montrent aussi que le classificateur KNN (K plus proche) a permis les meilleurs taux de classification (80,29% pour la position de la poche et 88% pour la position de la ceinture) comparé aux arbres de décision et au classificateur bayésien.

Cette étude représente quelques limites telles que la taille de la base de données qui est petite puisque le nombre de sujets était restreint à sept participants ce qui est limité pour la validation des résultats. Pour cela, nous suggérons d'augmenter la taille de la base de données en y ajoutant de nouvelles données pour de nouveaux sujets, et ce, pour tester et valider les différents classificateurs utilisés lors de cette étude.

Cette étude démontre qu'il est possible d'utiliser les appareils mobiles pour discriminer les différentes activités faites par un sujet ce qui permettra d'améliorer la qualité de vie des personnes surtout les personnes âgées au terme de sécurité et de télésurveillance de ces personnes. Il suffit pour cela d'améliorer l'application Android pour répondre aux besoins attendus, que ce soit au niveau sécuritaire ou santé.

D'autres axes de recherches pourraient intéresser des chercheurs dans le domaine de classification automatique en utilisant d'autres dispositifs, à savoir, les montres intelligentes ou le textile intelligent, puisqu'ils contiennent, eux aussi, des capteurs tels que les accéléromètres. L'application Android utilisée lors de cette étude pourra aussi être compatible pour d'autres systèmes (iOS et windows mobile) et il est possible de rendre cette application plus compatible et plus répondante aux besoins des futurs utilisateurs selon les activités physiques souhaitées.

# RÉFÉRENCES

## BIBLIOGRAPHIQUES

- [1] Samuel Ambapour « Introduction à l'analyse des données » document de travail pour le bureau d'application des méthodes statistiques et informatiques, 2003
- [2] Fenelon, J.-P., Qu'est-ce que l'analyse des données ? : exposé accessible aux non mathématiciens, Paris, Lefonen, 1981, 311 p.
- [3] Jean-Paul Benzécri. « Histoire et Préhistoire de l'Analyse des données » Livre Partie 5 » 1977,p.9-40.
- [5] Arnaud Martin. « Analyse des données » polycopié de cours ENSIETA – Réf :1463 Septembre 2004.
- [6] I. Jolliffe, Principal Component Analysis, 2nd edition éd., Springer-Verlag,2002.
- [7] F.-G. Carpentier. « Analyses multidimensionnelles et applications informatiques »Cours 2010-2011.
- [8] Youssef Ouakrim. " Classification de sujets asymptomatiques et gonarthrosique en fonction des données cinématiques : Comparaison de l'approche globale et de l'approche locale", thèse de maîtrise ÉTS 2011
- [9] [https://fr.wikipedia.org/wiki/Classification\\_automatique](https://fr.wikipedia.org/wiki/Classification_automatique), Consulté le 20/02/2016
- [10] Charles Bouveyron « Modélisation et classification des données de grande dimension. application à l'analyse d'images », thèse Université Joseph Fourier – Grenoble 2006.



- [11] Ludovic L, Alain M, Marie P. " Statistique exploratoire multidimensionnelle", Dunod, Paris, 1995
- [12] Roux, M. (1994) Classification des données d'enquête in Grangé, D. & Lebart, L., ed. Traitement statistique des enquêtes. Paris, Dunod, p. 91-112.
- [13] Rapkin, Bruce D; Luke, Douglas A (1993). Cluster Analysis in Community Research:  
Epistemology and Practice. American Journal of Community Psychology, 21 (2), 247-277.
- [14] Falissard, B. (2005). Méthodes de classification : Analyse en clusters in Comprendre et utiliser les statistiques dans les sciences de la vie. Paris: Masson. p. 275-296.
- [15] Falissard, B. (2005). Méthodes de segmentation, CART in Comprendre et utiliser les statistiques dans les sciences de la vie. Paris : Masson. p. 297-302.
- [16] Andes, N. (1998) Robust strategies for application of cluster analysis, Bulletin de Méthodologie Sociologique, Mars, no. 58, pp.31-47.
- [17] Blashfield, R.K., et M.S. Aldenderfer, (1998) The methods and problems of cluster analysis, In J.R. Nesselroade & R.B. Cattell (Eds.), The handbook of multivariate experimental psychology (2nd ed., pp.447-474). New-York: Plenum Press.
- [18] Everitt, B.S., Landau, S. AND Leese, M. (2001). Cluster Analysis. Fourth Edition. London: Arnold. 237p.
- [19] Milligan, G. W. Et M. Cooper., (1987) Methodology Review: Clustering Methods, tiré de Applied Psychological Measurement , 11, 329-354.
- [20] Nakache, J.P. et Confais, J. (2000). Méthodes de classification, Montreuil, France: Cisia-Ceresta, 185p.
- [21] Petrucci, A et S. Schifini-D'Andrea (2002). Objective and Subjective Indicators: A Spatial Analysis Using Classification Techniques, Social Indicators Research, 60, 55-88.

- [22] Dominique F (2010), Machine à vecteurs de support-Une introduction, CA MUS1, 7-25.
- [22] E. Lebarbier, T. Mary-Huard, Classification non supervisée, AgroParisTech 2008
- [23] Francis B, K-means, EM, Mélanges de Gaussiennes, Théorie des graphes. Octobre 2010
- [24] Marie C et Patrick L, Algorithme de Kohonen : classification et analyse exploratoire des données. Université de Paris-CNRS UMR 8595
- [25] Robertas D, Mindaugas V, Justas S, Marcin W, " Human activity recognition in AAL environments using random projections", Hindawi Publishing Corporation, Volume 2016, Article ID 4073584, 17 pages.
- [26] L. Atallah, B. Lo, R. King, and G.-Z. Yang, "Sensor positioning for activity recognition using wearable accelerometers," IEEE Transactions on Biomedical Circuits and Systems, vol. 5, no. 4, pp. 320–329, 2011.
- [27] A. Bayat, M. Pomplun, and D.A. Tran, "A study on human activity recognition using accelerometer data from smartphones," Procedia Computer Science, vol. 34, pp. 450–457, 2014.
- [28] M. Berchtold, M. Budde, D. Gordon, H. R. Schmidtke, and M. Beigl, "ActiServ: activity recognition service for mobile phones," in Proceedings of the 14th IEEE International Symposium on Wearable Computers (ISWC '10), pp. 1–8, IEEE, Seoul, South Korea, October 2010.
- [29] N. A. Capela, E. D. Lemaire, and N. Baddour, "Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients," PLoS ONE, vol. 10, no. 4, Article ID e0124414, 2015.
- [30] P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single triaxial accelerometer," IEEE Transactions on Biomedical Engineering, vol. 61, no. 6, pp. 1780–1786, 2014.

- [31] A. Henpraserttae, S. Thiemjarus, and S. Marukatat, "Accurate activity recognition using a mobile phone regardless of device orientation and location," in Proceedings of the International Conference on Body Sensor Networks (BSN '11), pp. 41–46, IEEE, Dallas, Tex, USA, May 2011.
- [32] E. Hoque and J. Stankovic, "AALO: activity recognition in smart homes using Active Learning in the presence of Overlapped activities," in Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare and Workshops (PervasiveHealth '12), pp. 139–146, May 2012.
- [33] T. Iso and K. Yamazaki, "Gait analyzer based on a cell phone with a single three-axis accelerometer," in Proceedings of the 8th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '06), pp. 141–144, Espoo, Finland, September 2006.
- [34] M. Kose, O. D. Incel, and C. Ersoy, "Online human activity recognition on smart phones," in Proceedings of the Workshop on Mobile Sensing: From Smartphones and Wearables to BigData (Colocated with IPSN), pp. 11–15, Beijing, China, April 2012.
- [35] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," ACM SIGKDD Explorations Newsletter, vol. 12, no. 2, pp. 74–82, 2011.
- [36] N. Lane, M. Mohammad, M. Lin et al., "Bewell: a smartphone application to monitor, model and promote wellbeing," in Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare, pp. 23–26, IEEE, 2012.
- [37] Y. S. Lee and S. Cho, "Activity recognition using hierarchical hidden markov models on a smartphone with 3d accelerometer," in Hybrid Artificial Intelligent Systems: 6th International Conference, HAIS 2011, Wroclaw, Poland, May 23–25, 2011, Proceedings, Part I, vol. 6678 of Lecture Notes in Computer Science, pp. 460–467, Springer, Berlin, Germany, 2011.

- [38] A. Mannini and A.M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010.
- [39] M. J. Mathie, B.G. Celler, N.H. Lovell, and A.C.F. Coster, "Classification of basic daily movements using a triaxial accelerometer," *Medical and Biological Engineering and Computing*, vol. 42, no. 5, pp. 679–687, 2004.
- [40] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN '06)*, pp. 113–116, IEEE, Cambridge, Mass, USA, April 2006.
- [41] E. Miluzzo, N. D. Lane, K. Fodor et al., "Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application," in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08)*, pp. 337–350, Raleigh, NC, USA, November 2008.
- [42] J. Parkkila, M. Ermes, P. Korhonen, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 119–128, 2006.
- [43] T. S. Saponas, J. Lester, J. Froehlich, J. Fogarty, and J. Landay, *iLearn on the iPhone: Real-Time Human Activity Classification on Commodity Mobile Phones*, 2008.
- [44] P. Siirtola and J. Rönning, "Recognizing human activities user independently on smartphones based on accelerometer data," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 1, no. 5, pp. 38–45, 2012.
- [45] T. Sohn, A. Varshavsky, A. LaMarca et al., "Mobility detection using everyday GSM traces," in *UbiComp 2006: Ubiquitous Computing: 8th International Conference, UbiComp 2006 Orange County, CA, USA, September 17–21, 2006 Proceedings*, vol. 4206 of *Lecture Notes in Computer Science*, pp. 212–224, Springer, Berlin, Germany, 2006.

- [46] J. Yang, "Toward physical activity diary: motion recognition using simple acceleration features with mobile phones," Proceedings of the 1st International Workshop on Interactive Multimedia for Consumer Electronics (IMCE '09), 2009.
- [47] C. Zhu and W. Sheng, "Motion- and location-based online human daily activity recognition," *Pervasive and Mobile Computing*, vol. 7, no. 2, pp. 256–269, 2011.
- [48] Oscar. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," in *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192-1209, Third Quarter 2013.
- [49] Andrea M, Angelo M S "Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers", *Sensors* 2010, 10(2), 1154-1175
- [50] Yi He and Ye Li, "Physical activity recognition utilizing the built-in kinematic sensors of a smartphone", Hindawi Publishing Corporation, Volume 2013, Article ID 481580, 10 pages.
- [51] Organisation mondiale de la santé, mHealth – New horizons for health through mobile technologies, Global Observatory for eHealth series – Volume 3.